REMARKS

Claims 1-11 are pending in the application. Claims 1-11 are rejected. All rejections are respectfully traversed.

The invention extracts speech recognition features from a speech signal coded as a bitstream. The bitstream is decoded to recover linear predictive coding filter parameters and to recover a residual signal. The linear predictive coding filter parameters and the residual signal are discriminatively combined into speech recognition features.

Claims 1 and 6 are rejected under 35 U.S.C. 102(b) as being anticipated by Hershkovits (U.S. Patent No. 6,003,004).

Hershkovits ***convolves*** a residual signal and LPC parameters in a short-term synthesis filter, item 86, Fig. 9, to produce a ***voice signal*** from an input compressed voice signal, i.e., LAR_{cr} data. The invention ***discriminatively combines*** LPC parameters and a residual signal into ***speech recognition features***. As would readily be understood by a person of ordinary skill in the art, synthesis filters, such as item 86, convolve two signals. A convolution is an integral that expresses the amount of overlap of one function *g* as it is shifted over another function *f*. It therefore "blends" one function with another. Depending on the spectra used for the convolution, the convolution operation can be an addition or a multiplication, as is known in the art. The output is a voice signal. That is exactly what decoders such as the one depicted in Figure 9 do. The process depicted by Figure 9 is exactly described at col. 7, lines 56-64, below:

FIG. 9 shows that the decoder includes an RPE decoder **80**, a long term predictor **84**, a short term synthesis filter **86**, and a de-emphasizer **88**. The RPE decoder **80** receives the M_{cr}, x_{maxcr} and x_{mcr} signals and generates a remnant signal e_r'. The long term predictor **84** uses the b_{cr} and N_{cr} signals to generate a residual signal d_r' from the remnant signal e_r'. The short term synthesis filter **86** generates the voice signal from the residual signal d_r' and the short term LPC parameters, transmitted in the form of the LAR_{cr} data.

Lines 62-64 clearly describe a synthesis filter convolving a residual signal and LPC parameters to produce a voice signal. MPEP 2131 explicitly states that in order to anticipate a claim "each and every element as set forth in the claims" must be found in the prior art reference." The identical invention must be shown in as complete detail as is contained in the ... claim." The Examiner's assertion that "the phrase "discriminatively combining" cited in the rejected claims is broad enough for the prior art reference to read on" is absurd, because a person of ordinary skill in the art would immediately understand the distinction between convolving and discriminatively combining. Discriminative combining takes features, in this case from the LPC parameters and the residual signal, stacks them in a single vector and performs some matrix operation on the vector. Examples given in the application include applying Fisher's linear discriminant analysis (LDA), or using a discriminatory neural network. Further, the discriminatory combining as claimed produces speech recognition features. Hershkovits' synthesis filter produces a voice signal. The two are totally different things.

Hershkovits makes it perfectly clear that LPC coefficients alone are used to generate speech recognition features, see col. 8, lines 50-53, below:

> As mentioned hereinabove with respect to FIG. **6**, once the LPC coefficients are extracted, they are transformed (step **70**) into the recognition features which the recognizer/ training step requires.

Hershkovits never describes combining anything with linear predictive coding filter parameters to produce speech recognition features, as claimed. Further, Hershkovits never describes discriminatory combining as understood in the art. Therefore, Hershkovits can never anticipate what is claimed and therefore, the Applicants respectfully request the Examiner reconsider and withdraw his rejection based on Hershkovits.

Regarding claim 6, Hershkovits only computes the *energy* of the residual signal. Claimed is analyzing an entire *spectrum* of the residual signal. One of ordinary skill in the art would never confuse 'energy', which we all know to mean power, with spectrum, which relates to signal frequency.

Furthermore, with all due respect, the Examiner's comparison makes no sense. A frame is 5 to 20 msec worth of samples. A frame was and never will be a spectrum. Disclosing the processing of frame samples does not anticipate analyzing the spectrum of a residual signal as claimed.

Claims 2-4 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hershkovits in view of Aguilar (U.S. Patent No. 6,691,082).

Claim 2 recites up-sampling the linear predictive coding parameters and interpolating the up-sampled linear predictive coding parameters. In claim 3, a set of samples is obtained for every frame of the bitstream. In claim 4, cepstral vectors

are derived from the up-sampled LPC filter parameters.

Aguilar upsamples the raw speech signal from 4 KHz to 8 KHz, not the LPC parameters themselves as claimed. Note also that the claimed cepstral vectors are obtained from *upsampled LPC parameters*, and not from an upsampled acoustic signal as in the recited references. Therefore, Aguilar can never be used to make the invention obvious.

Claim 5 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hershkovits in view of Park (U.S. Patent No. 6,108,624).

Park pads positions of a time axis corresponding to a second subframe with zeros, initializes the pitch filter and LPC filter to zero. Claimed are setting *short-term prediction coefficients* to zero. Short-term prediction coefficients are not positions on a time axis, nor pitch or LPC filters.

Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hershkovits in view of applicant's admitted prior art.

The application merely states that a 32-dimensional log spectra is derived from the residual signal. The specification at page 11 does not say "residual log-spectra **must** be derived before inputting into the neural network," this is an inaccurate characterization of the invention by the Examiner. The prior art cited at page 11 only has to do with speech recognition, generally. There is no admission or indication that the prior art teaches "deriving a high-dimensional log spectra from *up-sampled LPC parameters*," as stated in claim 7. The Examiner is requested to consider all limitations in the claim. As stated above with respect to claim 2,

Applicants believe that upsampling LPC parameters is novel. The prior art upsamples acoustic signals, and then derives LPC parameters.

Claims 8-10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hershkovits in view of applicant's admitted prior art, and further in view of Kuhn (U.S. Patent No. 6,343,267).

At column 9, Kuhn discloses:

> FIG. **5** shows how the maximum likelihood technique works. The input speech from the new speaker is used to construct supervector **70**. As explained above, the supervector comprises a concatenated list of speech parameters, corresponding to cepstral coefficients or the like. In the illustrated embodiment these parameters are floating point numbers representing the Gaussian means extracted from the set of Hidden Markov Models corresponding to the new speaker. Other HMM parameters may also be used. In the illustration these HMM means are shown as dots, as at **72**. When fully populated with data, supervector **70** would contain floating point numbers for each of the HMM means, corresponding to each of the sound units represented by the HMM models. For illustration purposes it is assumed here that the parameters for phoneme "ah" are present but parameters for phoneme "iy" are missing.
>
> The eigenspace **38** is represented by a set of eigenvectors

Specifically, Kuhn concatenates speech parameters such as cepstral coefficients with themselves. In contrast, the invention concatenates cepstral vectors with high-dimensional log spectra.

The claimed invention reduces the dimensionality of an extended vector that is a concatenation of a cepstral vector and a high-dimensional log-spectra derived from a residual signal. Applicants firmly believe this is novel. The Examiner has failed to show, and Applicants are unaware of, any prior art that describes this

combination of limitations. The reference cited at page 10 and 11 are unrelated to what is specifically claimed.

Claim 11 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hershkovits.

It should be noted that the application makes it clear that "the invention enables the design of a distributed speech recognition system where feature extraction need not be performed on a user's handheld device. This reduces the immediate to change existing coding and transmission standards in telephone networks. It should also be understood, the invention makes the type of codec used transparent to the speech recognizer, which is not the case when the features are extracted from a reconstructed bitstream."

It is well known that the typical distributed system is in terms of a client/server model. In the instant application, the client is a handheld communications device, such as a cell phone, and the server is operated by the service provider, e.g., the telephone company. In this scenario, it is desired to simplify the cell phone, and have the speech recognition done at the server. Up to now to now, devices that perform speech recognition do both the feature extraction and the recognition based on the extracted feature. In contrast, the invention does the extraction at the client, the cell phone, and the recognition itself at the server. It is this unexpected division of labor that provides advantages to applications designed according to the invention, particularly in a system with a distributed architecture.

All rejections have been complied with, and applicant respectfully submits that the application is now in condition for allowance. The applicant urges the Examiner to

contact the applicant's attorney at the phone and address indicated below if assistance is required to move the present application to allowance. Please charge any shortages in fees in connection with this filing to Deposit Account 50-0749.

Respectfully Submitted,

Andrew J. Curtin
Registration No. 48,485

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, 8th Floor
Cambridge, MA 02139
Telephone: (617) 621-7573
Facsimile: (617) 621-7550